

1. はじめに

深層学習では、事前学習したモデルを転移先のデータセットを用いたファインチューニングにより、短期間で高い精度を実現できる。このとき、どのようなデータセットで事前学習するかが重要である。また、より高い精度を実現するには、複数のモデルを用いて共同学習する知識転移グラフ [1] がある。知識転移グラフは、知識の伝達方法を自動最適化することにより、様々な学習方法を表現できる。この知識転移グラフにおいても事前学習は重要であると考えられる。そこで、本研究では知識転移グラフにおける事前学習の有効性について調査する。

2. 知識転移グラフ

知識転移グラフ [1] は、ネットワークモデルをノード、ノード間の知識転移をエッジとしてグラフ構造で共同学習を表現している。ノードは精度を高める対象となるターゲットノードとターゲットノードの学習をサポートする補助ノードに分類される。また、エッジは損失を制御する4種のゲート関数で表される。ゲート関数には、値に変化を加えない Through gate, 常に0を出力する Cutoff gate, 学習時間と共に出力が増加する Linear gate, 正解した場合にのみ損失を出力する Correct gate があり、これらを組み合わせることで様々な知識転移を実現することができる。知識転移グラフのハイパーパラメータはモデルとゲート関数であり、これらを自動最適化することにより最適な知識転移法を獲得する。

3. 知識転移グラフにおける事前学習

知識転移グラフにおける事前学習の有効性を確認する。図1に知識転移グラフの構造を示し、以下に詳細な手順を示す。

- Step1 事前学習用データセットを用いて単体学習
 Step2 各ノードには、Step1で学習した重み、もしくは乱数で初期化
 Step3 ファインチューニング用データセットを用いて知識転移グラフを探索

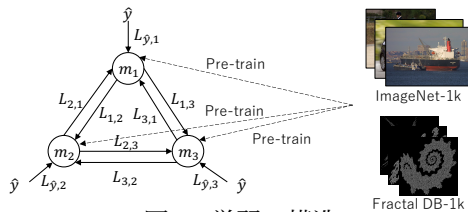


図1: 学習の構造

事前学習には ImageNet-1k または Fractal DataBase-1k (Fractal DB-1k) を使用する。ImageNet-1k は 120 万枚の大規模な画像データセットであり、Fractal DB-1k はフラクタル画像で構成された 100 万枚の画像データセットである。最適化におけるハイパーパラメータは以下の通りとし、グラフ探索の最適化手法は ASHA を使用する。

ノード：事前学習を用いないモデル，事前学習したモデル
 エッジ：Through gate, Cutoff gate, Linear gate, Correct gate

4. 実験

知識転移グラフにおける事前学習の有効性を確認するために、各ノードにおいて事前学習の有無による精度評価を行い、各ノードの事前学習の有無をハイパーパラメータに含めた知識転移グラフの探索を行う。

4.1. 実験概要

探索や評価用のデータセットとして、12,554 枚で構成される自動車画像識別データセットを使用する。最適化手法は SGD, 学習回数は 200 エポック, 初期学習率は 0.01, バッチサイズは 64 である。ネットワークモデルは ResNet18 を使用する。知識転移グラフのノード数は 3 とし、ノード 1 をターゲットノードとする。

事前学習の有無による評価実験では、全てのエッジを知識伝達に手を加えない Through gate のグラフを使用する。知識転移グラフの探索の実験では、事前学習なしのみ、ImageNet-1k 事前学習の有無、Fractal DB-1k 事前学習の有無、ImageNet-1k 事前学習と Fractal DB-1k 事前学習の混在において、最適なグラフ構造を 1,500 回探索する。

4.2. 事前学習の有無による精度評価

表1に実験結果を示す。精度は3回試行した時の平均精度と標準偏差である。表1から、ターゲットノードであるノード1に事前学習を取り入れることで精度が向上することが確認できる。また、全てのノードで事前学習を用いない場合と比較して、ImageNet-1k で事前学習すると最大 4.92pt 向上し、Fractal DB-1k で事前学習すると最大で 3.82pt 向上した。この結果から、条件を固定した知識転移グラフにおいて事前学習は有効であると考えられる。

表1: 各ノードの事前学習の有無による精度評価

事前学習			精度 [%]	
ノード 1	ノード 2	ノード 3	ImageNet-1k	Fractal DB-1k
			90.45 ± 0.42	
	✓		92.06 ± 0.30	91.97 ± 0.99
		✓	92.25 ± 0.50	92.20 ± 0.24
	✓	✓	93.13 ± 0.30	93.84 ± 0.61
✓			94.26 ± 0.21	93.46 ± 0.18
✓	✓		95.11 ± 0.43	94.08 ± 0.64
✓		✓	95.07 ± 0.30	94.27 ± 0.03
✓	✓	✓	95.37 ± 0.25	93.96 ± 0.08

4.3. 事前学習の有無における最適グラフの探索

精度を向上させるためのグラフ構造を探索する。図2に最適化したグラフ構造を示す。ここで、RN18 は未学習の ResNet18 であり、PIRN18 は ImageNet-1k で事前学習した ResNet18 であり、PFRN18 は Fractal DB-1k で事前学習した ResNet18 である。図2の赤の楕円で表されたターゲットノードであるノード1の精度を比較すると、事前学習なしよりも ImageNet-1k 事前学習で 4.43pt, Fractal DB-1k 事前学習で 3.41pt, 事前学習混在で 4.64pt 高精度なグラフ構造を獲得した。この結果から、知識転移グラフにおいて異なる事前学習モデルを用いることが有効であるといえる。

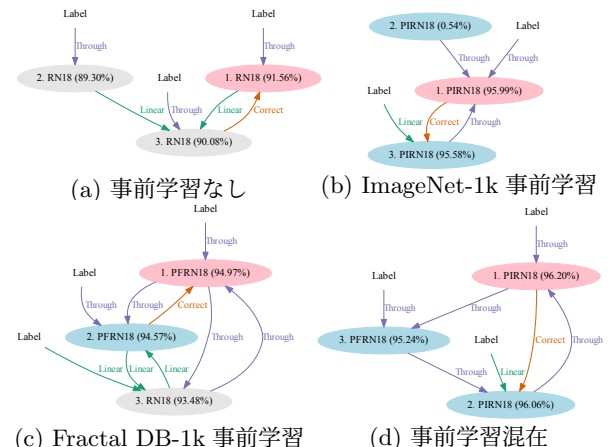


図2: 最適化により獲得した知識転移グラフ

5. おわりに

本稿では、知識転移グラフにおける事前学習の有効性を調査した。実験結果から、知識転移グラフにおいて事前学習は有効性を確認した。今後は、他の事前学習用データセットに変更し探索することで、さらに精度が高くなるようなグラフ構造が獲得可能であるかを調査する。

参考文献

- [1] S. Minami, *et al.*, “Knowledge Transfer Graph for Deep Collaborative Learning”, ACCV, 2020.